

III

ANALYSES

Contents	III-1
Maps of Papillomavirus mRNA Transcripts	III-3
<i>Carl Baker and Charles Calef</i>	
Transcription Factor Binding Sites in the LCR of Genital HPVs	III-21
<i>Mark O'Connor, Shih-Yen Chan, and Hans-Ulrich Bernard</i>	
A Closer Look at Papillomavirus Variants	III-41
<i>Cossette Wheeler and Joseph Icenogle</i>	
The E6 Protein	III-47
<i>Gerald Myers and Elliot Androphy</i>	
HPV16 E7: Primary Structure and Biological Properties	III-58
<i>Aaron Halpern and Karl Münger</i>	
Analysis of the Sequences of the L1 and L2 Capsid Proteins of Papillomaviruses	III-74
<i>Joseph Icenogle</i>	
HPV Patterns and Their Similarities	III-91
<i>Gerald Myers and Andrew Farmer</i>	
Assessing Recombination in HPV, Part I	III-124
<i>Aaron Halpern</i>	
Assessing Recombination in HPV, Part II	III-139
<i>Andrew Farmer and Gerald Myers</i>	

Maps of Papillomavirus mRNA Transcripts

Carl Baker^a and Charles Calef^b

^a *Laboratory for Tumor Virus Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892-5055*

^b *MS K710, Los Alamos National Laboratory, Los Alamos, New Mexico 87545*

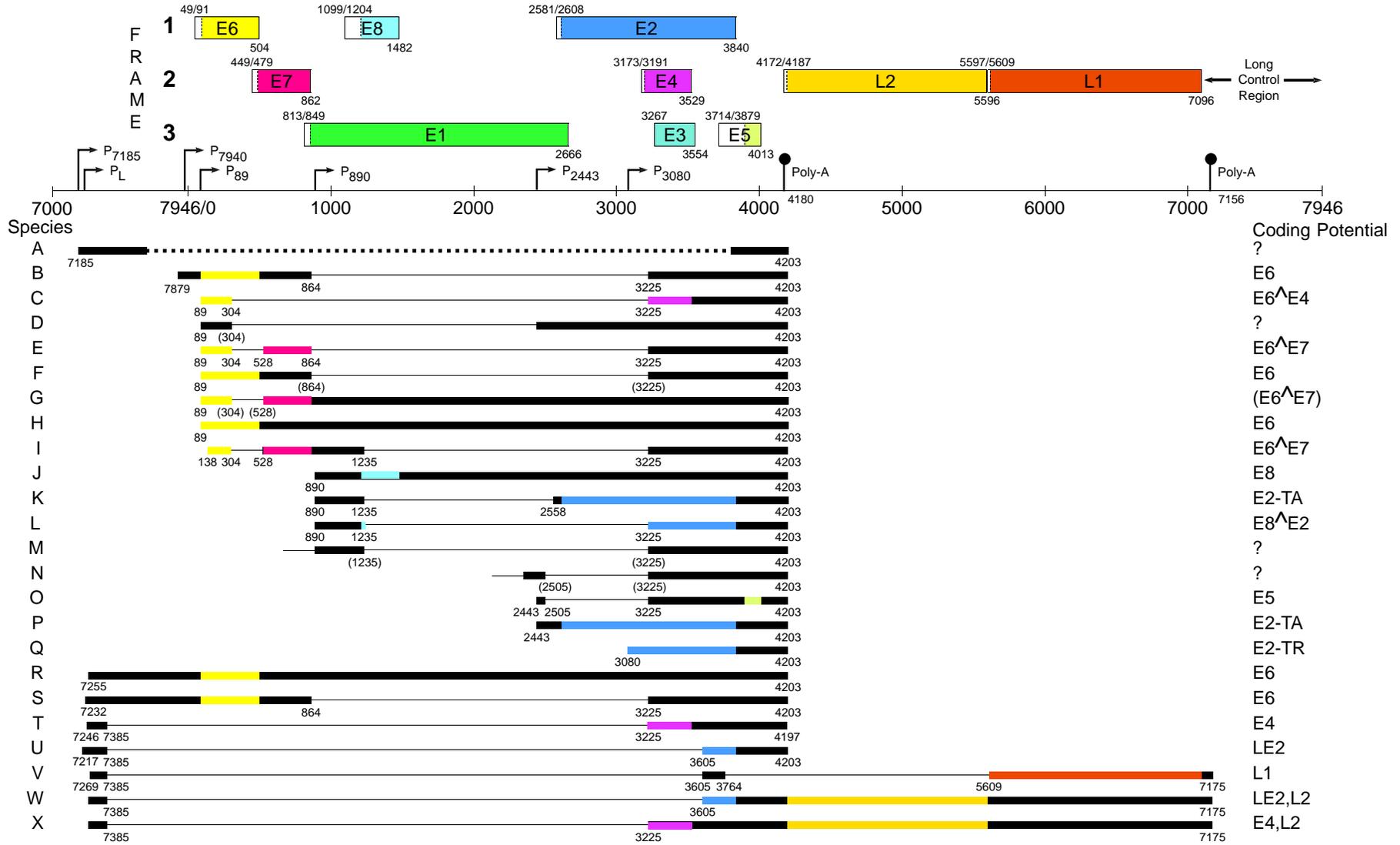
Explanation of Maps

Maps showing the open reading frames and mRNA transcripts of seven papillomaviruses are presented on the following pages. Facing each map is a brief description of the transcripts, and following the maps is a list of references from which the maps were compiled. In each map the significant ORFs are shown in their proper reading frames as colored rectangles. At the upper left end of the rectangle are two numbers. The first corresponds to the nucleotide (nt) position of the ORF start, the first nucleotide following a stop codon. The second number records the nt position of the first ATG, which is also indicated by a dotted line within the rectangle. The position of the last nt in the stop codon of each ORF is printed at the lower right corner of the rectangles. The rectangle's colored fill begins at the ATG and extends to the end; thus it corresponds to the coding sequence, not the entire ORF. In the cases where no ATG exists in the ORF, only one number is present in the upper left corner and the rectangle is completely filled with color. Below the ORFs is a scale of the genome divided into thousands. On the scale are placed the positions of promoters (represented by arrows) and the poly(A) signals. The exact position of the poly(A) signal is printed below the scale line. Located below the genome scale are diagrams of mRNA species, most of which are spliced. The exons are illustrated by heavy black lines, while the introns are indicated by black hairlines between. The numbers printed below the lines indicate the 5' and 3' termini of the RNAs, and the 5' and 3' splice junction positions. The splice junction numbers give the position of the last nucleotide in the exon before the splice and the position of the first nucleotide of the exon following the splice. Splice junctions in parentheses were deduced from the genomic sequence and have not been confirmed by cDNA sequencing. Where 5' or 3' ends of the RNAs are uncertain, no nt position is given. Superimposed on the exons are colored rectangles representing the gene, or part thereof, coded by that portion of the exon. The coding potential of each transcript is also listed at the right. In that list a ^ symbol between two gene name (e.g. E1^E4) indicates a fusion product. The * symbol indicates different forms of the E6 product.

BPV-1 has served as the prototype for the genetic analysis of the papillomaviruses. The mRNAs from BPV-1 transformed C127 cells as well as productively infected bovine fibropapillomas have been exhaustively analyzed. The top of the BPV-1 map shows the genomic organization of BPV-1. The long control region is labelled and, in order to show the position of promoters, is repeated to the left of nt 0. The locations of known promoters [3,6,23,24,27] are indicated by arrows and labeled P_n where n is the approximate nucleotide position of the RNA start site for that promoter. P_{7185} , P_{89} , P_{890} , P_{2443} , and P_{3080} are also referred to as P_1 , P_2 , P_3 , P_4 , and P_5 [6,23], respectively. The late promoter (P_L) is the major promoter active in the fibropapilloma and is the only promoter not active in C127 cells transformed by BPV-1 [3]. The early and late polyadenylation signals are at nt 4180 and 7156, respectively. The structures of BPV-1 mRNAs from BPV-1 transformed mouse C127 cells (species A—Q) were determined by cDNA cloning as well as electron microscopy, nuclease protection, PCR, and primer extension. The 5' most ORF containing a translation initiation codon and a significant coding region is indicated at the right of each mRNA. Although an E6/E7 fusion ORF is the 5' most ORF for species I, the cDNA from which this structure was deduced has been shown to encode the E1 M protein [26].

Additional very rare mRNA species from cycloheximide treated BPV-1 transformed C127 cells have been characterized [6], but are not shown here. The structures of mRNAs unique to the BPV-1 fibropapilloma (species R—X) were determined by RT-PCR and cDNA cloning and sequencing [3,4]. Although the E2 and E4 ORFs are the first significant ORF for species W and X, these mRNAs may also encode the L2 protein. A more detailed discussion of BPV-1 transcription can be found elsewhere, including references for each mRNA [1,2].

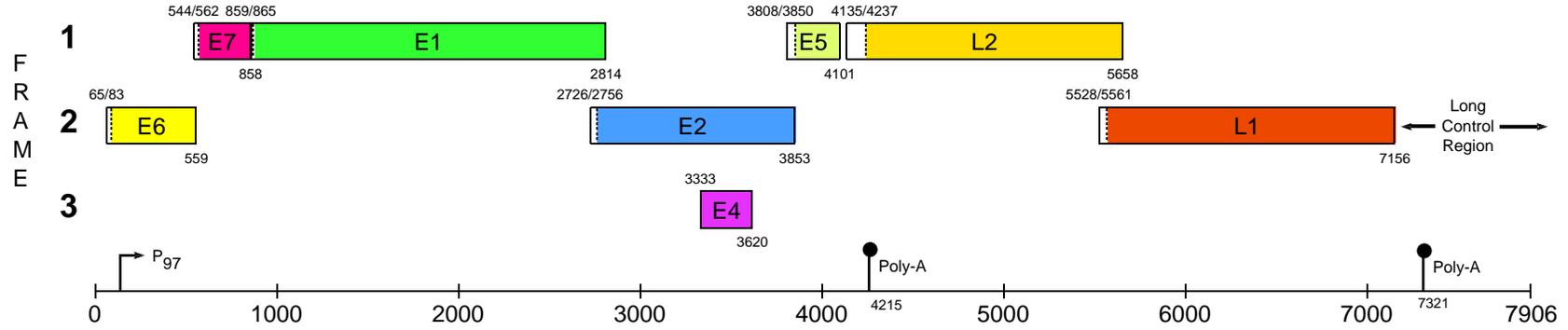
BPV1R



III-5
OCT 95

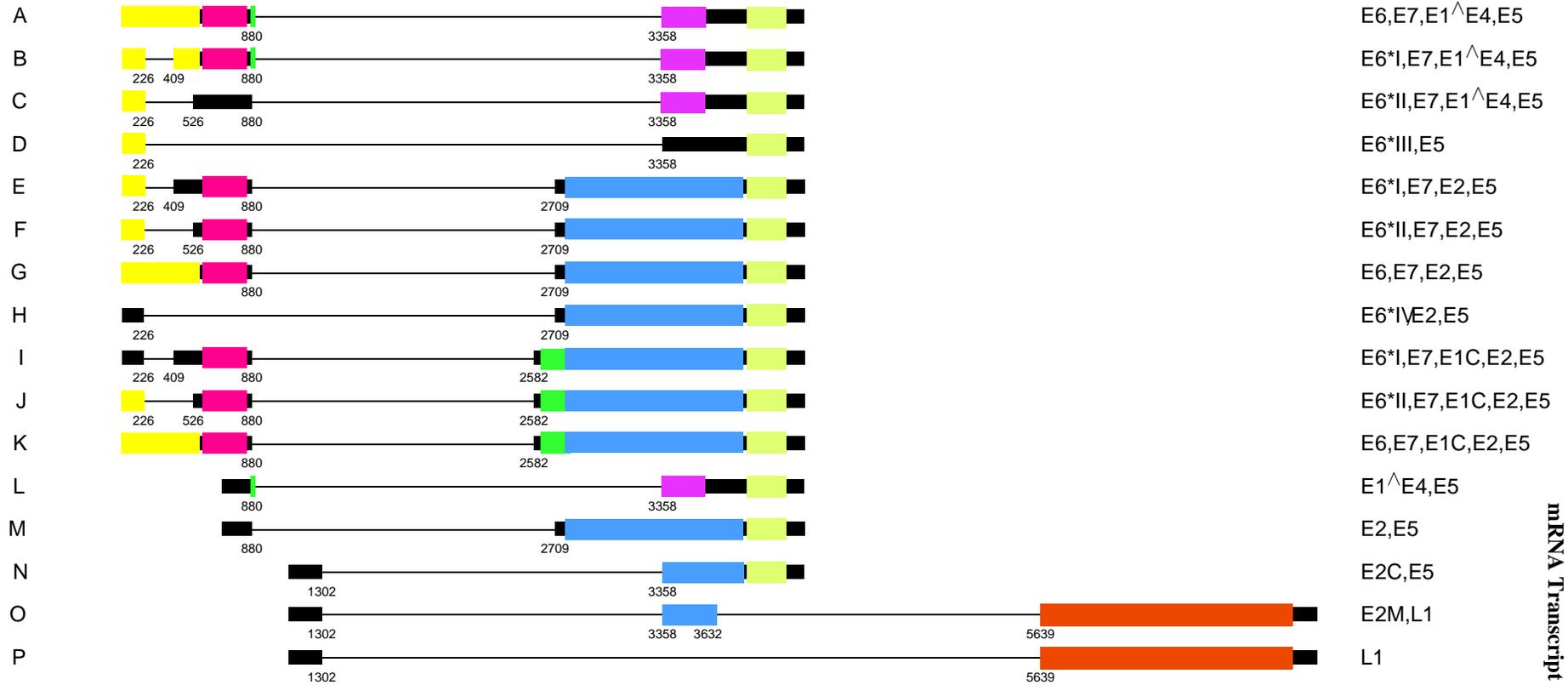
HPV-16 mRNAs isolated from transfected cells and a variety of tumor cell lines and lesions containing both extrachromosomal and integrated HPV-16 genomes have been analyzed in multiple laboratories [8,15,18-20]. Viral/host chimeric mRNAs have been purposely omitted. All nucleotide positions correspond to the revised HPV-16 sequence published in Part I of this compendium. Most mRNA species were determined by RT-PCR, so the identity of the 5' and 3' ends are not known. However, only one promoter (P₉₇) has been definitively identified for HPV-16 [22]. Species A-K are most likely transcribed from this promoter. The promoter responsible for the transcription of the late mRNAs (species O-P) is not known. The early and late polyadenylation signals are located at nt 4215 and nt 7321, respectively [11,15].

HPV16R



Species

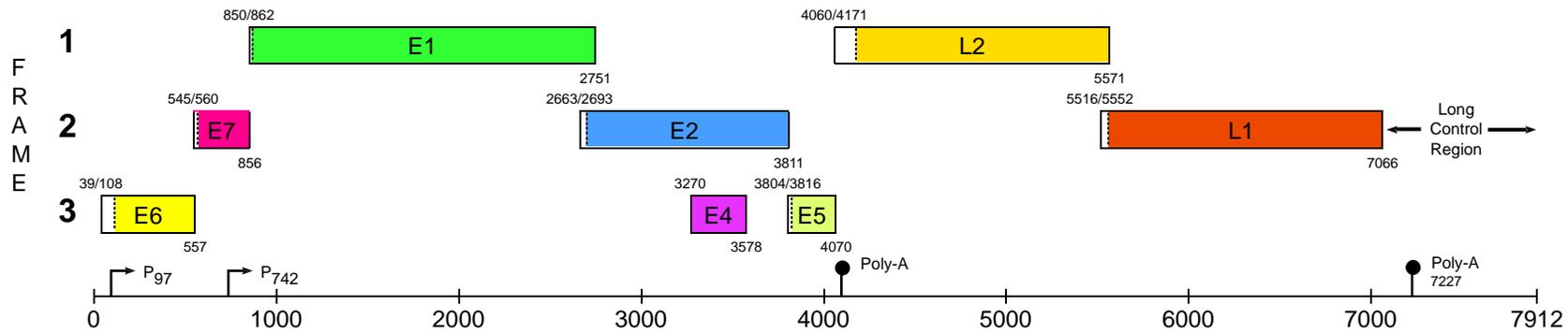
III-7
OCT 95



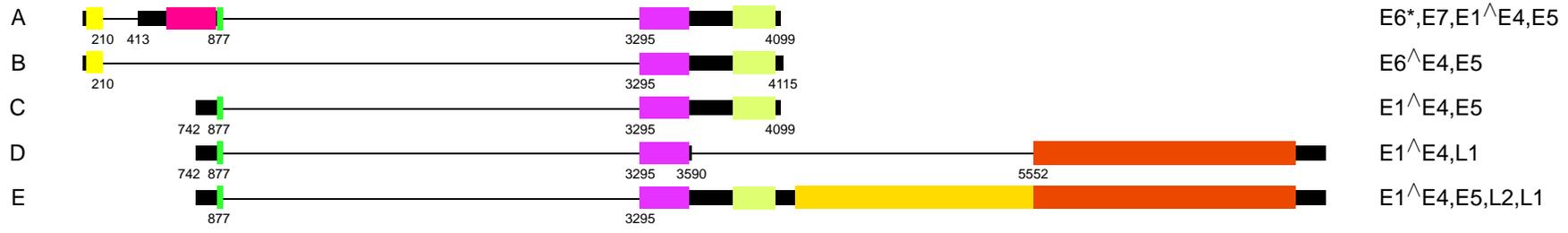
mRNA Transcript Maps

HPV-31 mRNAs were investigated in CIN62 cells grown in monolayer cultures and in organotypic raft cultures [9,14]. CIN612 cells contain extrachromosomal HPV-31. Two promoters were identified by primer extension and RNase protection experiments. P₉₇ is active in both monolayer and raft cultures; P₇₄₂ is active only in differentiated raft cultures [9]. No promoter could be identified in the E6 ORF [9]. The mRNA structures identified are shown in the facing map. Species A and B are presumably transcribed from P₉₇. Species C, D, and presumably E are transcribed from P₇₄₂. Additional L1 mRNAs appear to be transcribed from P₉₇, but the exact splice structure of these mRNAs is unknown. Early mRNAs were polyadenylated between nt 4099 and 4125 [9]. The exact sites of polyadenylation at the late poly(A) site were not determined in these studies, but a putative polyadenylation signal exists at nt 7227.

HPV31



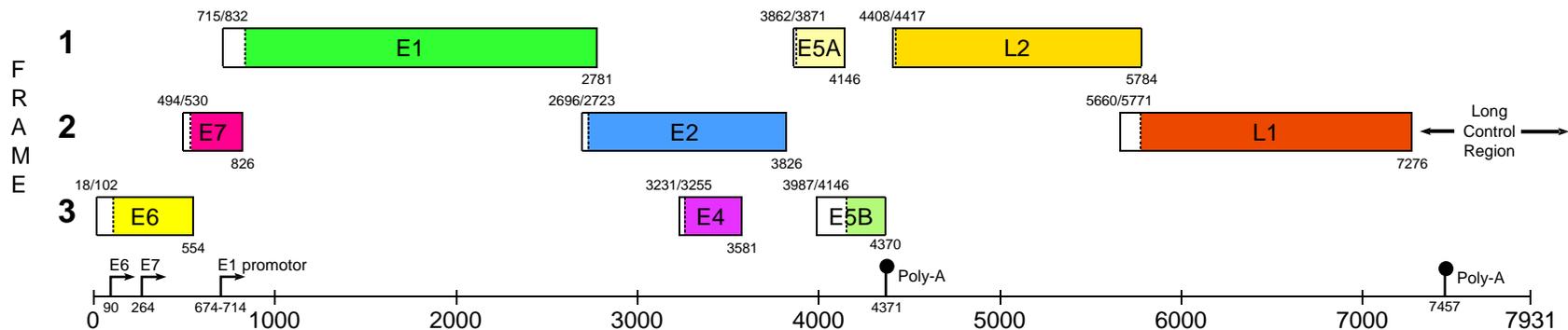
Species



III-9
OCT 95

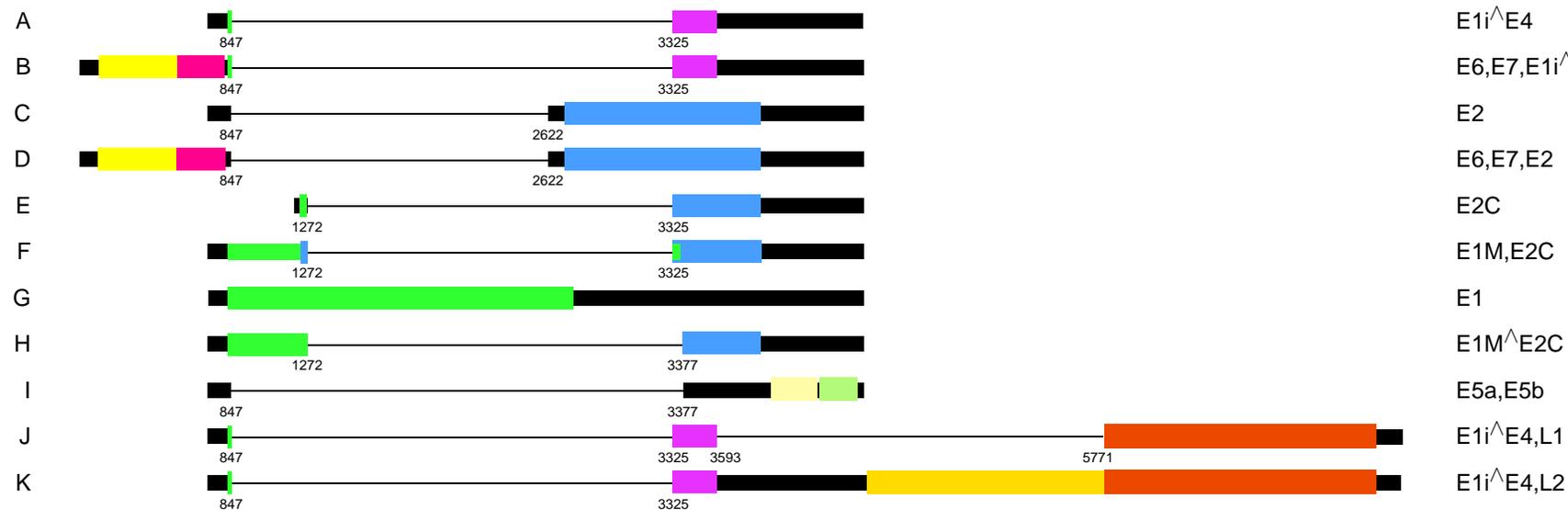
HPV-11 mRNAs, isolated from genital condyloma acuminata and experimental condylomatous cysts implanted under the renal capules of nude mice, have been analyzed by several laboratories [5,7,13,16,17]. The types of analyses include cDNA cloning (species A), electron microscopy analysis of R-loops (species A–G, J–K), and RT-PCR analysis (species C–F, H–J). The deduced mRNA structures are shown in the HPV-11 map. The 5' ends of individual mRNAs have not been mapped at the nucleotide level. However, three promoters have been mapped by Smotkin et al. [21] using nuclease S1 analysis: the E6 promoter initiates at nt 90, the E7 promoter at nt 264, and the E1 promoter at nts 674–714. The 5' end of the cDNA representing species A is located at nt 716 [13], so this mRNA is presumably transcribed from the E1 promoter. The early poly(A) site has been identified from analysis of species A cDNAs [13]. The late poly(A) site has not been identified experimentally, but the L1 and L2 mRNAs are presumably polyadenylated utilizing the poly(A) signal at nt 7457.

HPV11



III-11
OCT 95

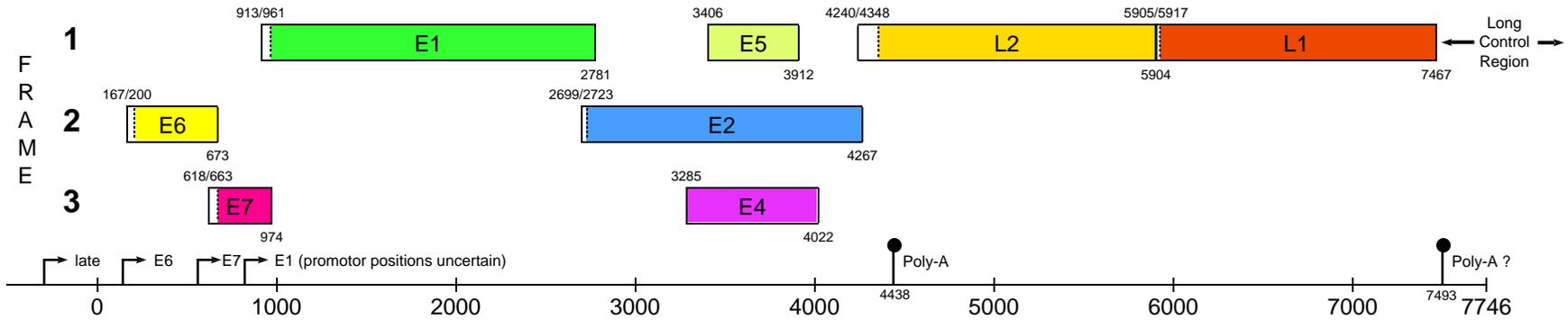
Species



mRNA Transcript Maps

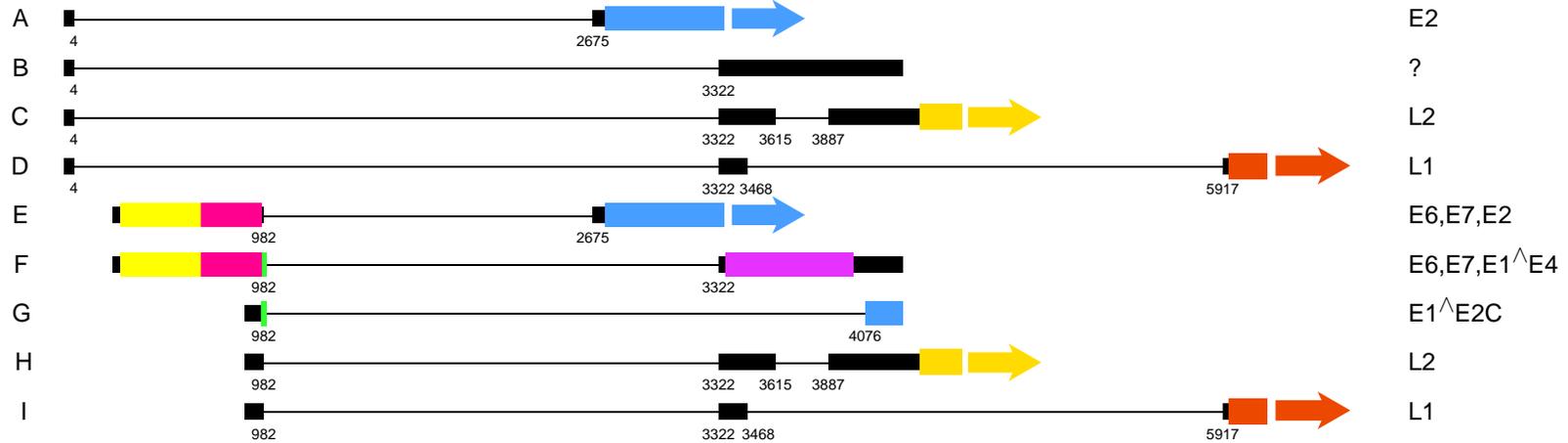
HPV-5 mRNAs isolated from benign skin lesions from patients with epidermodysplasia verruciformis were analyzed by RT-PCR (Haller, Stubenrauch, and Pfister, *Virology*, in press). The 5' and 3' ends of these mRNAs are unknown. However, in situ hybridization data as well as promoter mapping data for the related virus HPV-8 suggest that four promoters may exist, a late promoter in the LCR, an E6 promoter, an E7 promoter, and an E1 promoter similar to that identified for HPV-31 [9]. Species A–D are most likely transcribed by the late promoter and species E–F by the E6 promoter. The early and late poly(A) sites have not been experimentally determined. However, by analogy with the related virus HPV-47 [12], the early poly(A) signal is at nt 4438. Species A–B and E–G are most likely polyadenylated at the early poly(A) site, while species C–D and H–I at the late poly(A) site.

HPV5



Species

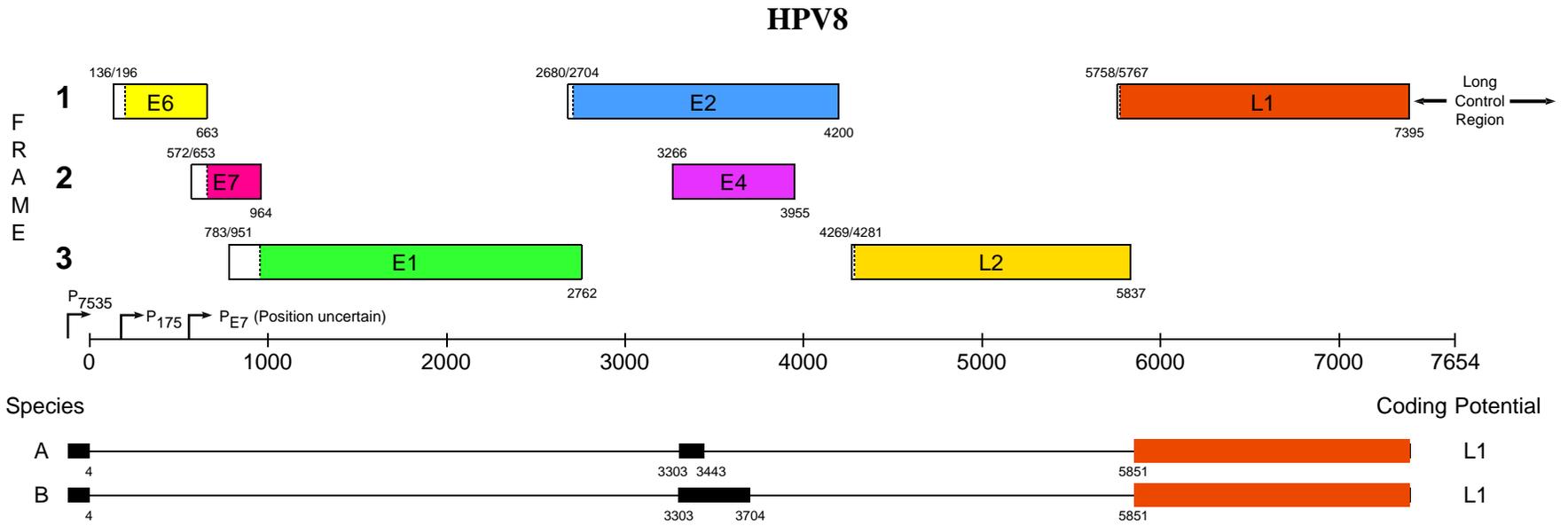
Coding Potential



III-13
OCT 95

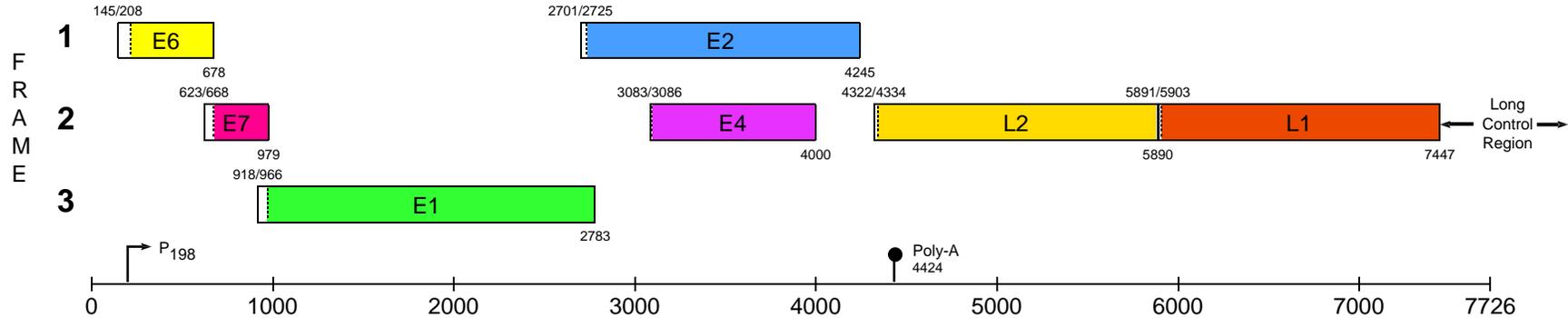
HPV-8 mRNAs have been investigated from both transfected cells and lesions from an epidermodysplasia verruciformis patient. Analysis of RNA start sites suggest at least three promoters: a late promoter in the LCR with start site at nt 7535 (P₇₅₃₅) [25], an E6 promoter with start sites at nt 175–179 (P₁₇₅) [10,25], and an E7 promoter with start site just upstream of the E7 ORF (not further mapped [10]). RT-PCR analysis of RNA from the EV lesion [25] identified two L1 mRNAs which presumably are transcribed from the late promoter. The exact 5' and 3' ends of these mRNAs have not been determined, however.

III-15
OCT'95



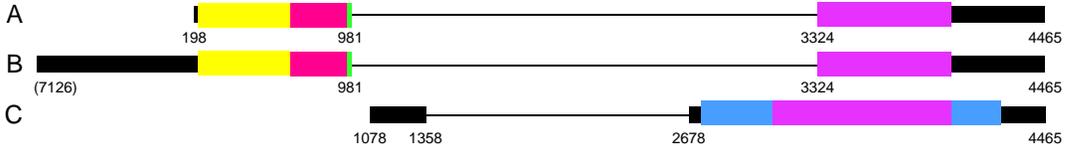
Analysis of mRNA from an HPV-47 transformed rat cell line using S1 and primer extension analysis revealed RNA start sites at nt 183–202 with a major start site at nt 198 [12]. The early poly(A) sites were mapped to nt 4444 and nt 4465 by cDNA cloning and sequencing. Three RNA structures were determined using cDNA cloning coupled with S1 analysis. Species A is the most abundant mRNA. Only one clone was identified for species B and C mRNAs. The 5' end of the cDNA insert for the species B clone was at an EcoRI site which was used for cloning the cDNA, so the actual cDNA may have extended further upstream.

HPV47



III-17
OCT 95

Species



References

- 1 Baker, C. C. 1990. Bovine papillomavirus type 1 transcription, p. 91–112. In H. Pfister (ed.), *Papillomavirus and human cancer*. CRC Press, Inc. Boca Raton, FL.
- 2 Baker, C. C. 1993. The genomes of the papillomaviruses, p. 1.134–1.146. In S. O'Brien (ed.), *Genetic maps: locus maps of complex genomes*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- 3 Baker, C. C. and P. M. Howley. 1987. Differential promoter utilization by the bovine papillomavirus in transformed cells and productively infected wart tissues. *EMBO J.* **6**:1027–1035.
- 4 Barksdale, S. K. and C. C. Baker. 1993. Differentiation-specific expression from the bovine papillomavirus type 1 P2443 and late promoters. *J. Virol.* **67**:5605–5616.
- 5 Chiang, C.-M., T. R. Broker, and L. T. Chow. 1991. An E1ME2C fusion protein encoded by human papillomavirus type 11 is a sequence-specific transcription repressor. *J. Virol.* **65**:3317–3329.
- 6 Choe, J., P. Vaillancourt, A. Stenlund, and M. Botchan. 1989. Bovine papillomavirus type 1 encodes two forms of a transcriptional repressor: Structural and functional analysis of new viral cDNAs. *J. Virol.* **63**:1743–1755.
- 7 Chow, L. T., M. Nasser, S. M. Wolinsky, and T. R. Broker. 1987. Human papillomavirus types 6 and 11 mRNAs from genital condylomata acuminata. *J. Virol.* **61**:2581–2588.
- 8 Dootbar, J., A. Parton, K. Hartley, L. Banks, T. Crook, M. Stanley, and L. Crawford. 1990. Detection of novel splicing patterns in a HPV16-containing keratinocyte cell line. *Virology* **178**:254–262.
- 9 Hummel, M., J. B. Hudson, and L. A. Laimins. 1992. Differentiation-induced and constitutive transcription of human papillomavirus type 31b in cell lines containing viral episomes. *J. Virol.* **66**:6070–6080.
- 10 Itiner, T., G. Sagner, H. Pfister, and F. O. Wetstein. 1990. The E7 protein of human papillomavirus 8 is a nonphosphorylated protein of 17 kDa and can be generated by two different mechanisms. *Virology* **179**:428–436.
- 11 Kennedy, I. M., J. K. Haddow, and J. B. Clements. 1990. Analysis of human papillomavirus type 16 late mRNA 3' processing signals in vitro and in vivo. *J. Virol.* **64**:1825–1829.
- 12 Kiyono, T., K. Nagashima, and M. Ishibashi. 1989. The primary structure of major viral RNA in a rat cell line transfected with type 47 human papillomavirus DNA and the transforming activity of its cDNA and E6 gene. *Virology* **173**:551–565.
- 13 Nasser, M., R. Hirochika, T. R. Broker, and L. T. Chow. 1987. A human papilloma virus type 11 transcript encoding an E1-E4 protein. *Virology* **159**:433–439.
- 14 Pray, T. R. and L. A. Laimins. 1995. Differentiation-dependent expression of E1 E4 proteins in cell lines maintaining episomes of human papillomavirus type 31b. *Virology* **206**:679–685.

- 15 Rohlf, M., S. Winkenschach, S. Meyer, T. Rupp, and M. Durst. 1991. Viral transcription in human keratinocyte cell lines immortalized by human papillomavirus type-16. *Virology* **183**:331–342.
- 16 Rotenberg, M. O., C.-M. Chiang, M. L. Ho, T. R. Broker, and L. T. Chow. 1989. Characterization of cDNAs of spliced HPV-11 E2 mRNA and other HPV mRNAs recovered via retrovirus-mediated gene transfer. *Virology* **172**:468–477.
- 17 Rotenberg, M. O., L. T. Chow, and T. R. Broker. 1989. Characterization of rare human papillomavirus type 11 mRNAs coding for regulatory and structural proteins, using the polymerase chain reaction. *Virology* **172**:489–497.
- 18 Sherman, L. and N. Alloul. 1992. Human papillomavirus type 16 expresses a variety of alternatively spliced mRNAs putatively encoding the E2 protein. *Virology* **191**:953–959.
- 19 Sherman, L., N. Alloul, I. Golan, M. Durst, and A. Baram. 1992. Expression and splicing patterns of human papillomavirus type-16 mRNAs in pre-cancerous lesions and carcinomas of the cervix, in human keratinocytes immortalized by HPV 16, and in cell lines established from cervical cancers. *Int. J. Cancer* **50**:356–364.
- 20 Smitis, H. L., M. T. E. Cornelissen, M. F. Jebbink, J. G. Van den Tweel, A. P. H. B. Struyk, M. Briet, and J. ter Schegget. 1991. Human papillomavirus type 16 transcripts expressed from viral-cellular junctions and full-length viral copies in CaSki cells and in a cervical carcinoma. *Virology* **182**:870–873.
- 21 Smotkin, D., H. Prokoph, and F. O. Wettstein. 1989. Oncogenic and nononcogenic human genital papillomaviruses generate the E7 mRNA by different mechanisms. *J. Virol.* **63**:1441–1447.
- 22 Smotkin, D. and F. O. Wettstein. 1986. Transcription of human papillomavirus type 16 early genes in a cervical cancer and a cancer-derived cell line and identification of the E7 protein. *Proc. Natl. Acad. Sci. USA* **83**:4680–4684.
- 23 Stenlund, A., G. L. Brean, and M. R. Botchan. 1987. A promoter with an internal regulatory domain is part of the origin of replication in BPV-1. *Science* **236**:1666–1671.
- 24 Stenlund, A., J. Zabielski, H. Ahola, J. Moreno-Lopez, and U. Pettersson. 1985. Messenger RNAs from the transforming region of bovine papilloma virus type I. *J. Mol. Bio.* **182**:541–554.
- 25 Stubenrauch, F., J. Malejczyk, P. G. Fuchs, and H. Pfister. 1992. Late promoter of human papillomavirus type 8 and its regulation. *J. Virol.* **66**:3485–3493.
- 26 Thorner, L., N. Bucay, J. Choe, and M. Botchan. 1988. The product of the bovine papillomavirus type I modulator gene (M) is a phosphoprotein. *J. Virol.* **62**:2474–2482.
- 27 Yang, Y. C., H. Okayama, and P. M. Howley. 1985. Bovine papillomavirus contains multiple transforming genes. *Proc. Natl. Acad. Sci. USA* **82**:1030–1034.

mRNA Transcript Maps